

WHAT IS CLAIMED IS:

1. A method comprising:
 segmenting video data to create a video clip based
 on timing data; and
 determining a most likely gesture in the video clip.

2. The method of claim 1, wherein determining includes
 determining a probability that each of a plurality of
 predefined gestures are performed in the video clip contains
 the predefined gesture.

3. The method of claim 2, wherein determining the
 probability that the video clip contains each of the
 predefined gesture includes evaluations of Hidden Markov
 Models.

4. The method of claim 1, wherein the timing data
 includes beat data corresponding to a beat of audio data.

5. The method of claim 4, further comprising:
 receiving the audio data; and
 extracting the beat data from the audio data.

6. The method of claim 4, wherein the video clip
 includes a portion of the video data corresponding to a
 predefined time window surrounding the occurrence of at least
 one beat.

1 7. The method of claim 1, further comprising displaying
2 a target gesture to be performed by the subject of the video
3 data.

1 8. The method of claim 1, wherein each video clip
2 contains video frames.

1 9. The method of claim 1, further comprising
2 identifying moving regions in each video frame in the video
3 clip.

1 10. The method of claim 9, further comprising generating
2 a feature vector for each video frame of the video clip.

1 11. The method of claim 1, further comprising generating
2 a score based on whether the video clip contains the target
3 gesture.

1 12. The method of claim 11, further comprising
2 displaying the score.

1 13. The method of claim 1, wherein determining if the
2 video clip contains the predefined gesture includes generating
3 a gesture probability vector having a plurality of elements,
4 each element being associated with one of a plurality of
5 predefined gestures and representing a probability that the

6 video clip contains each of the associated predefined
7 gestures.

1 14. A system comprising:
2 a temporal segmentor connected to receive video data and
3 to create a video clip from the video data based on timing
4 data; and
5 a recognition engine, in communication with the temporal
6 segmentor, to determine if the video clip contains a
7 predefined gesture.

1 15. The system of claim 14, wherein the recognition
2 engine includes a plurality of Hidden Markov Models.

1 16. The system of claim 14, further comprising:
2 a timing data source, in communication with the temporal
3 segmentor, to provide the timing data to the temporal
4 segmentor; and
5 a video source, in communication with the temporal
6 segmentor, to provide the video data to the temporal
7 segmentor.

1 17. The system of claim 14, further comprising a move
2 subsystem, in communication with the timing data source, to
3 provide a target gesture to be performed by the subject of the
4 video data.

1 18. The system of claim 17, wherein the target gesture
2 is a dance move that is to be performed by the subject of the
3 video data.

1 19. The system of claim 17, further comprising a scoring
2 subsystem, in communication with the recognition engine and
3 the move subsystem, to determine if the video clip contains
4 the target gesture.

1 20. The system of claim 19, further comprising a display
2 subsystem, in communication with the scoring subsystem, to
3 display a score that is a function of whether the video clip
4 contains the target gesture.

1 21. The system of claim 20, wherein the display
2 subsystem is in communication with the move subsystem and is
3 configured to display a gesture request based on the target
4 gesture.

1 22. The system of claim 14, wherein the recognition
2 engine is configured to recognize predefined gestures and to
3 produce a gesture probability vector having elements, each
4 element being associated with one of the predefined gestures
5 and representing the probability that the video clip contains
6 the associated predefined gesture.

1 23. The system of claim 14, wherein the timing data
2 source includes:

3 an audio source that provides an audio data; and
4 a beat extractor, in communication with the audio source,
5 that extracts beat data from the audio data.

1 24. The system of claim 23, wherein the video clip
2 corresponds to a beat in the beat data.

1 25. The system of claim 24, wherein the video clip
2 includes a portion of the video data corresponding to a
3 predefined time window surrounding the occurrence of the beat.

1 26. A computer program product, tangibly stored on a
2 computer-readable medium, for recognizing gestures contained
3 in video data, comprising instructions operable to cause a
4 programmable processor to:

5 segment the video data to create a video clip based
6 on timing data; and

7 determine if the video clip contains a predefined
8 gesture.

1 27. The product of claim 26, further comprising
2 instructions operable to cause the programmable processor to:
3 extract beat data from an audio signal; and

segment the video data to create the video clip using the beat data.

28. An audio-visual processing system including:
a video source to provide video data;
an audio source to provide audio data;
a speaker to play at least a portion of the audio data;
and
a computer program product, tangibly stored on a computer-readable medium, for recognizing gestures contained in video data, comprising instructions operable to cause a programmable processor, in communication with the video source and the audio source, to:

extract beat data from the audio data;
segment the video data to create a video clip based on beat data; and
determine if the video clip contains a predefined gesture.

29. The video processing system of claim 28, wherein the computer program product further includes instructions operable to cause the programmable processor to:
perform a Hidden Markov Model process to determine if the video clip contains the predefined gesture.

- 1
- 2
- 3